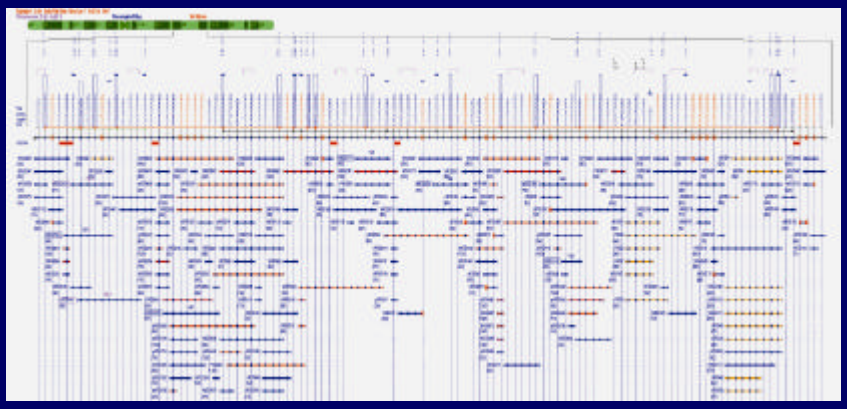
Physical Mapping



Eric D. Green, M.D., Ph.D.
Chief, Genome Technology Branch (NHGRI)
Director, NIH Intramural Sequencing Center (NISC)
Phone: 301-402-0201
FAX: 301-402-4735
E-Mail: egreen@nhgri.nih.gov

Physical Mapping: Outline

I. Fundamentals of Physical Mapping

II. Radiation Hybrid (RH) Mapping

III. Clone-based Physical Mapping

- A. Cloning Systems
- B. Strategies for Clone-based Physical Mapping
- C. Clone-based Physical Maps of Mammalian Genomes

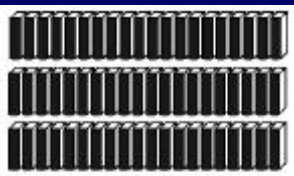
IV. Future Prospects

Physical Mapping: Goals

- Stress the Practical Aspects of Physical Mapping
- Focus More on the Mapping of Mammalian Genomes
- Highlight Relevant Literature
- Provide Information on Relevant Electronic Resources

Genome Sizes

Human Genome
Mouse Genome



~3,000,000,000 bp

Fruit Fly Genome

~160,000,000 bp

Nematode Genome

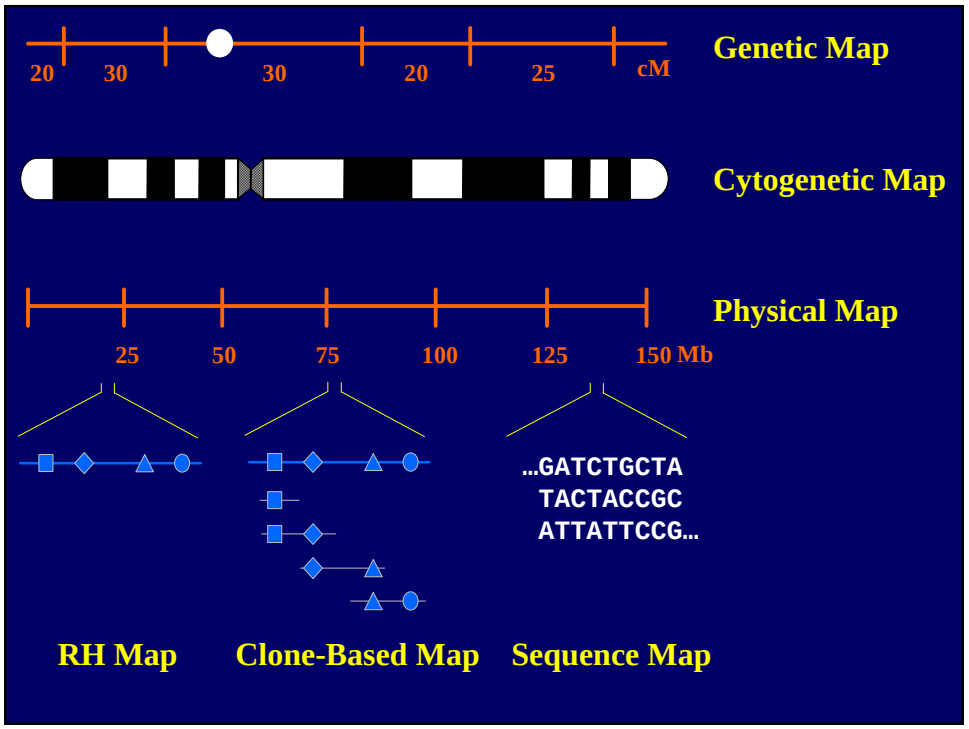
~100,000,000 bp

Yeast Genome

~15,000,000 bp

***E. coli* Genome**

~5,000,000 bp



Fundamentals of Physical Mapping

- Importance of Physical Maps:
  - Localization and Isolation of Genes (e.g., Positional Cloning)
  - Study of Genome Organization and Evolution
  - Framework for Systematic DNA Sequencing
- Mapping is About Order
- Physical Mapping Involves:
  - Ordering of Clones and/or Landmarks
  - Typically with Some Physically Measurable Metric
- General Types of Physical Maps:
  - Landmark Only
  - Clone-based
  - Sequence

Landmark Only Physical Maps

- Restriction Mapping by Pulsed-Field Gel Electrophoresis

  Riethman et al. (1997) *Genome Analysis*, Vol. 1, Chap. 2

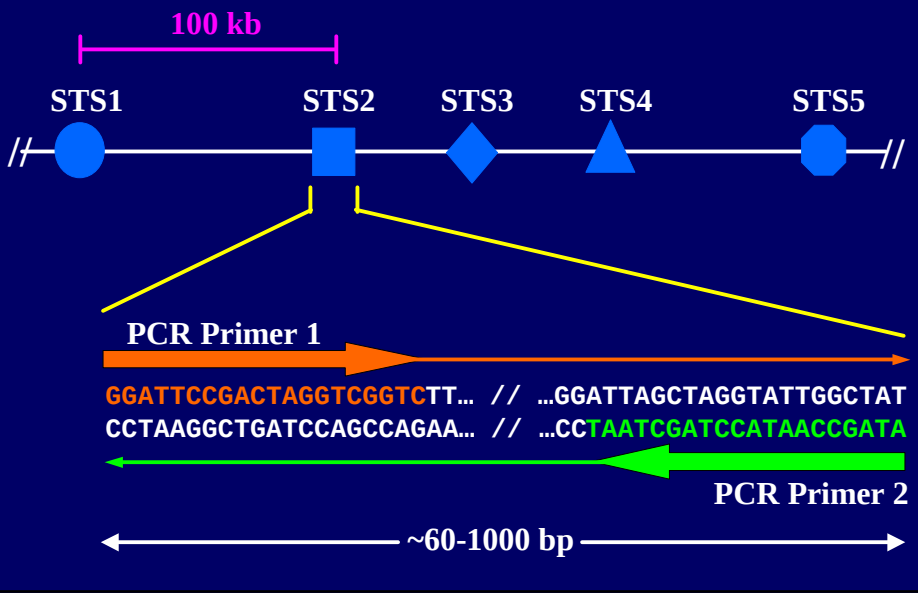
- Radiation Hybrid (RH) Mapping

  Matise et al. (1999) *Genome Analysis*, Vol. 4, Chap. 6

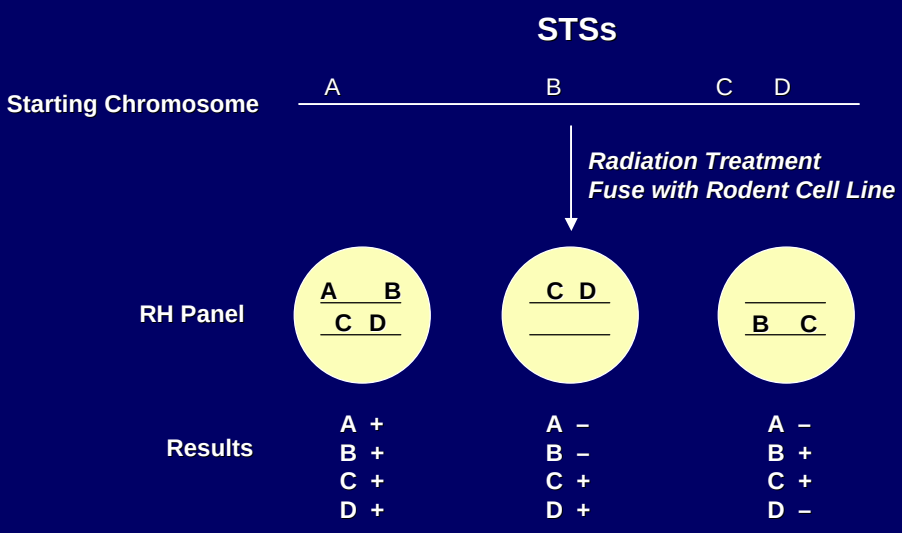
  Lecture by Dr. Tara Matise, *Current Topics '99*
  (see www.nhgri.nih.gov/COURSE99)

  Stanford University Genome Center
  (see shgc-www.stanford.edu)

Sequence-Tagged Sites (STSs)



RH Mapping: Theory



'Classic' Human RH Mapping Panels

	Genebridge 4 (GB4)	Stanford G3 (G3)
X-ray dosage	3,000 rad	10,000 rad
Map units	3,000 cR	10,000 cR
Cell lines	93	83
Average retention	32%	16%
Average fragment size	25 Mb	2.4 Mb
Effective resolution	1 Mb	0.25 Mb
Utility	*Long-range continuity*	*Higher resolution*

RH Mapping: Available Resources

· DNA from RH Panels

Research Genetics: www.resgen.com

·RH Mapping Servers

Human:

www.sanger.ac.uk/Software/RHserver/RHserver.shtml
www-shgc.stanford.edu/RH/G3index.html
carbon.wi.mit.edu:8000/cgi-bin/contig/rhmapper.pl

Mouse:

www.genome.wi.mit.edu/cgi-bin/mouse_rh/rhmapauto/rhmapper.cgi

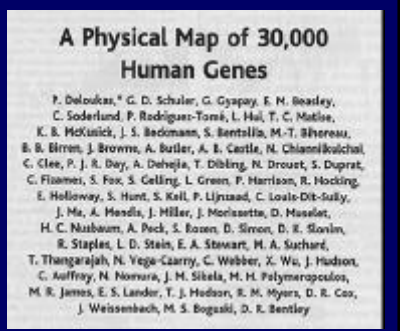
Rat:

rgd.mcw.edu/RHMAPSERVER/

·General Reference Information: compgen.rutgers.edu/rhmap

RH Mapping-based Human Gene Map

www.ncbi.nlm.nih.gov/genemap99


A Physical Map of 30,000 Human Genes

P. Deloukas,* G. D. Schuler, G. Gyapay, E. M. Beasley, C. Soderlund, P. Rodriguez-Tomé, L. Hui, T. C. Matise, K. B. McKusick, J. S. Beckmann, S. Bentolila, M.-T. Bihoreau, B. B. Birren, J. Browne, A. Butler, A. B. Castle, N. Chiannilkulchai, C. Clee, P. J. R. Day, A. Dehejia, T. Dibling, N. Drouot, S. Duprat, C. Fizames, S. Fox, S. Gelling, L. Green, P. Harrison, R. Hocking, E. Holloway, S. Hunt, S. Keil, P. Lijnzaad, C. Louis-Dit-Sully, J. Ma, A. Mondis, J. Miller, J. Morissette, D. Muselet, H. C. Nusbaum, A. Peck, S. Rozen, D. Simon, D. R. Slonim, R. Staples, L. D. Stein, E. A. Stewart, M. A. Suchard, T. Thangarajah, N. Vega-Czarny, C. Webber, X. Wu, J. Hudson, C. Auffray, N. Nomura, J. M. Sikela, M. H. Polymeropoulos, M. R. James, E. S. Lander, T. J. Hudson, R. M. Myers, D. R. Cox, J. Weissenbach, M. S. Boguski, D. R. Bentley

Science 282:744-746, 1998

Radiation hybrid map of the mouse genome

William J. Van Etten[1], Robert G. Steen[1], Huy Nguyen[1], Andrew B. Castle[1], Donna K. Slonim[1], Bing Ge[2], Chad Nusbaum[1], Greg D. Schuler[1], Eric S. Lander[1,4] & Thomas J. Hudson[1,2]

Nature Genetics
22:384-387 (1999)

A radiation hybrid map of the rat genome containing 5,255 markers

Takeshi K. Watanabe[1*], Marie-Therese Bihoreau[2*], Linda C. McCarthy[2,3*], Susanna E. Kiguwa[2*], Hiromaga Hirayasu[1], Atsushi Tsuji[1], Julie Browne[2], Yuki Yamasaki[1], Ayako Mizoguchi-Miyakita[1], Keiko Ogi[1], Toshihide Ono[1], Shiro Okuno[1], Naohide Kanemoto[1], Ei-ichi Takahashi[1], Kazuhiro Tomita[1], Hiromi Hayashi[1], Masakazu Adachi[1], Caleb Webber[2], Mario Davis[2], Susanne Rist[2], Catherine Knights[2], Angela Smith[2], Ricky Critcher[2], Jonathan Miller[2], Thiru Thangarajah[2], Philip J.R. Day[2], James R. Hudson Jr[4], Yasuo Irie[1], Toshihisa Takagi[5], Yusuke Nakamura[6], Peter N. Goodfellow[3], G. Mark Lathrop[2], Akira Tanigami[1] & Michael R. James[2]

Nature Genetics
22:27-36 (1999)

A High-Density Integrated Genetic Linkage and Radiation Hybrid Map of the Laboratory Rat

Robert G. Steen,[1,8] Anne E. Kwitek-Black,[2,8] Christopher Glenn,[3,8] Jo Gullings-Handley,[2] William Van Etten,[1] O. Scott Atkinson,[2] Diane Appel,[1] Simon Twigger,[2] Melanie Muir,[1] Tim Mull,[2] Mary Granados,[2] Mushira Kissebah,[1] Kerri Russo,[1] Robbin Crane,[1] Michael Popp,[1] Marc Peden,[2] Tara Matise,[4] Donna M. Brown,[3] Jian Lu,[2] Stephen Kingsmore,[1] Peter J. Tonellato,[2] Steve Rozen,[1] Donna Slonim,[1] Peter Young,[1] Margit Knoblauch,[2] Abraham Provoost,[2] Detlev Ganten,[2] Steven D. Colman,[1] Jonathan Rothberg,[2] Eric S. Lander,[1] and Howard J. Jacob,[2,9]

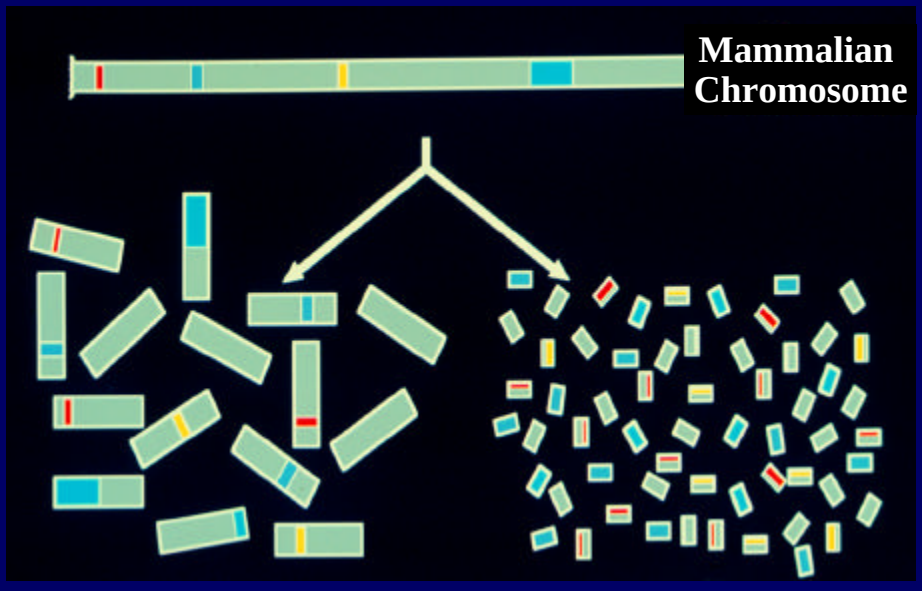
Genome Research
9:AP1-AP8 (1999)

A radiation hybrid map of the zebrafish genome

Robert Geisler[1], Gerd-Jörg Rauch[1], Herwig Baier[1], Fredrike van Bebber[1], Linda Bro[1], Marcus P.S. Dekens[1], Karin Finger[1], Cornelia Fritz[1], Michael A. Gates[1], Horst Geiger[1], Silke Geiger-Rudolph[1], Doreen Gilmour[1], Stefanie Glaser[1], Lara Gnügge[1], Hinrich Habeck[1], Katy Haffter[1], Scott Holley[1], Jeremy Keenan[1], Anette Kirn[1], Holger Knaut[1], Derek Lashkari[1], Florian Maderspacher[1], Ulrike Martyn[1], Stephan Neuhauss[1], Carl Neumann[1], Teresa Nicolson[1], Francisco Pelegri[1], Russell Ray[1], Jern M. Rick[1], Thierry Roehl[1], Tobias Roeser[1], Heike E. Schauerte[1], Alexander F. Schier[1], Ulrike Schönberger[1], Heiko-Lemm Schönthaler[1], Stefan Schulte-Merker[1], Catrin Seydler[1], William S. Talbot[1], Christina Weiler[1], Christiane Nüsslein-Volhard[1] & Pascal Haffter[1]

Nature Genetics
23:86-89 (1999)

Jigsaw Puzzle Analogy of Clone-based Physical Mapping



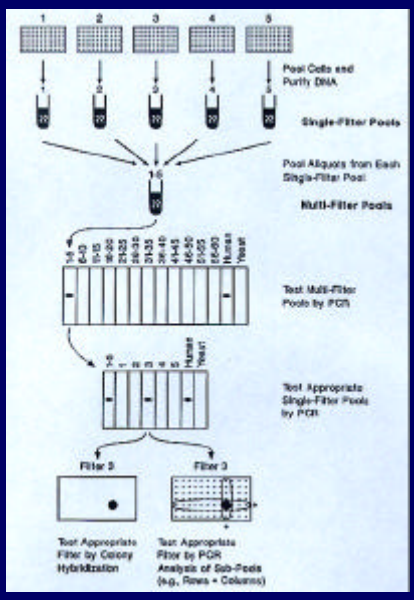
Clones for Physical Mapping: General Points

- Want Cloned DNA to Accurately Reflect the Starting Genome
  - Problem of Instability
  - Problem of Chimerism
- Development of 'Array Mentality' for Clone Libraries
  - Clones Arrayed in Individual Wells of Microtiter Plates
  - Various Densities (e.g., 96-and 384-Well Plates)
- Advantages of Arrayed Libraries ('Reference Libraries')
  1. Simplicity of Storing and Transferring Clone Collections
  2. Convenient Format for Retrieving Clones of Interest
  3. Ability to Assimilate Data on Common Clones
  4. Repeated PCR-based Screening
  5. Repeated Hybridization-based Screening

Commericial Involvement in Clone Distribution

Research Genetics:	www.resgen.com
Incyte Genomics:	www.incyte.com
ATCC:	www.atcc.org
BACPAC Resource:	www.chori.org/bacpac

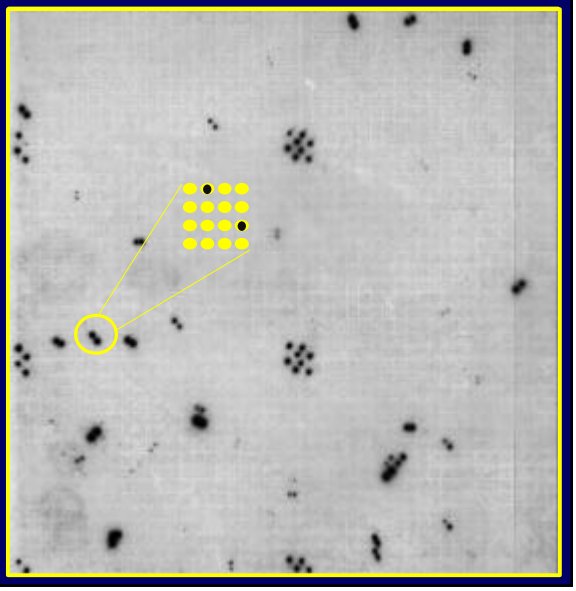
Screening Clone Libraries: PCR-based Approaches



Green & Olson, *PNAS* 87:1213-1217, 1990

Screening Clone Libraries: Hybridization-based Approaches

- 6 Fields, 16 x 384 BACs
- ~18,000 Unique Clones
- 4 x 4 Array
- Clones in Duplicate



Cosmids

- Bacterial-based Cloning System
- 'Antique' of the Large DNA Cloning Systems
- Plasmid Vector with Bacteriophage Packaging Sequences (*cos* Sites)
- High-Efficiency Packaging System
  - Relatively Homogeneous Insert Sizes
  - Libraries from Small Amounts of DNA (e.g., Flow-Sorted DNA)
  - Antibiotic Selection
- Cloned Inserts: 35-45 kb, Circular DNA
- High Copy Number
  - High Yields of DNA by Standard Methods
  - Instability Problems (Despite Recombination-Deficient Hosts)
- Relatively Non-Chimeric
- Various Libraries (Whole Genomes, Individual Chromosomes)
- References: Sambrook et al. (1989), Wahl et al. (1987), Ivens et al. (1993), Evans (1998)
- 'Fosmids' [Kim et al. (1992)]: Cosmid Vector Engineered with F Factor [Low Copy → More Stable]

P1 Clones

- Bacterial-based Cloning System
- Developed by Sternberg (1990)
- P1-based Vector and Complex P1 Packaging Extracts
  - Limited to 100 kb (Constraints of Viral Particle)
  - 2 loxP Sites Results in Circularization of DNA
  - Antibiotic Selection
- Cloned Inserts: 70-100 kb, Circular DNA
- Low Copy Number
  - Low Yields of DNA by Standard Methods
  - Highly Stable (with Recombination-Deficient Hosts)
  - Potential for IPTG Induction→ 10-30 Fold Increase
- Relatively Non-Chimeric
- Human and Mouse Libraries Commercially Available
- References: Sternberg (1990), Sternberg et al. (1990), Shepherd et al. (1994), Sternberg (1998)

P1-Derived Artificial Chromosomes (PACs)

- Bacterial-based Cloning System
- Developed by Ioannou et al. (1994)
- Slightly Modified P1 Vector
  - Lacks Packaging Signal
  - Antibiotic Selection
- Transform by Electroporation
- No Packaging of DNA → Larger Size Capacity
- Cloned Inserts: 100-150 kb, Circular DNA
- Low Copy Number
  - Low Yields of DNA by Standard Methods
  - Highly Stable (with Recombination-Deficient Hosts)
- Relatively Non-Chimeric

Bacterial Artificial Chromosomes (BACs)

- Bacterial-based Cloning System
- Developed by Shizuya et al. (1992)
- Based on the *E. coli* F Factor (Fertility Plasmid): Replication Control
- BAC Vectors
  - Cloning site in LacZ Gene (Blue/White Selection)
  - Antibiotic Selection
- Transform by Electroporation
- No Packaging of DNA → Larger Size Capacity
- Cloned Inserts: 100-200 kb, Circular DNA
- Low Copy Number
  - Low Yields of DNA by Standard Methods
  - Highly Stable (with Recombination-Deficient Hosts)
- Relatively Non-Chimeric
- Numerous Libraries Available (see www.chori.org/bacpac)
- See Birren et al. (1998)

Yeast Artificial Chromosomes (YACs)

- Yeast-based Cloning System (*Saccharomyces cerevisiae*)
- Developed by Burke et al. (1987)
- System Based on Ability to 'Harness' Cloned DNA with Structural Elements Required for the Propagation of a Linear Chromosome in Yeast
- Cloned Insert: ~100 kb to >1,000 kb, Linear DNA
- Spheroplast Transformation Procedure
  - Technically Demanding
  - Poorly Defined Upper Size Limit for Cloned Insert
- References: Hieter et al. (1990), Ramsay & Wicking (1991), Schlessinger & Kere (1992), Green et al. (1998)

Major Features of YACs

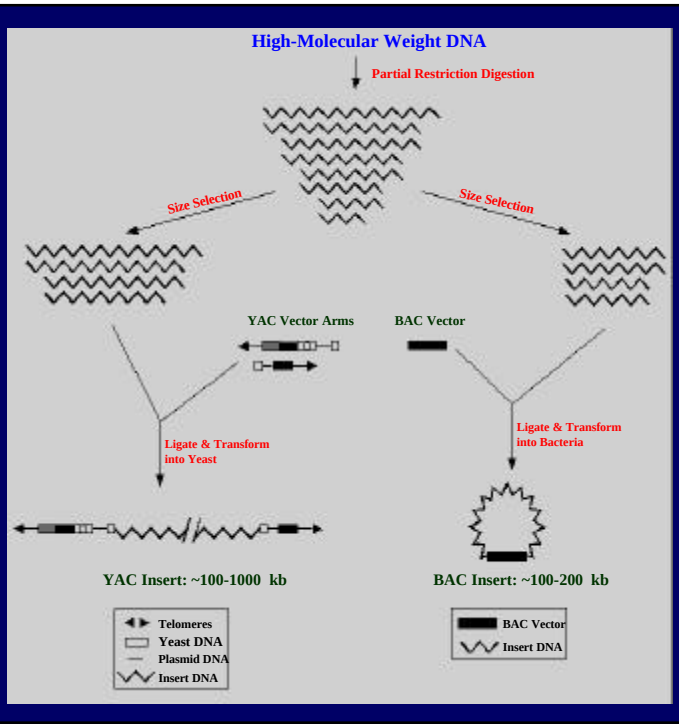
- Cloned DNA in Single Copy within Yeast Genome
  - Generally Same Structure and Size as Endogenous Chromosomes
  - Limited 'Access' to Cloned DNA (e.g., Gel Isolation)
- Chimerism as Major 'Problem' (Green et al., 1991)
  - Upwards of 40-60% of Clones in Total Mammalian DNA Libraries
- Instability (e.g., Internal Deletions) as Minor 'Problem'
- Various Human, Mouse, Rat, (and Other) Libraries Constructed

Human:
- Washington University [Burke and Olson (1991), Brownstein et al. (1989)]
- CEPH (Includes 'Mega-YACs') [Albertsen et al. (1990), Dausset et al. (1992)]
- ICRF [Larin et al. (1991)]
- ICI [Anand et al. (1989), Anand et al. (1990)]

Mouse:
- Princeton [Burke et al. (1991), Rossi et al. (1992)]
- St. Mary's [Chartier et al. (1992)]
- ICRF [Larin et al. (1991, 1993)]
- Whitehead [Kusumi et al. (1993), Haldi et al. (1996)]

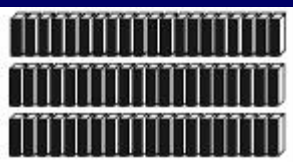
Rat:
- Harvard [Cai et al. (1997)]
- Whitehead [Haldi et al. (1997, 1997)]

Construction of Large-Insert Clones: YACs and BACs



Genome Sizes

Human
Mouse



~3,000,000,000 bp

Fruit Fly

~160,000,000 bp

Nematode

~100,000,000 bp

Yeast

~15,000,000 bp

E. coli

~5,000,000 bp

Cloning Capacity

YAC

~1,000,000 bp

BAC

~100,000 bp

Cosmid

~45,000 bp

Bacteriophage

~25,000 bp

Strategies for Clone-based Physical Mapping

- Two Key Components ('Jigsaw Puzzle Analogy')

  Cloned Fragments (Pieces of the Puzzle)
  Landmarks (Provide Clues for Aligning Pieces)

- Involves the Use of Landmarks to Assembly Clone 'Contigs'

  Contig: Overlapping Set of Clones that Together Contains a Contiguous Segment of the Source Genome

- Nature of Landmarks

  Must Provide 'Unique' Information About the DNA
  Must be Easy to Identify

- Early Candidates for Landmarks: Restriction Sites

  E. Coli [Kohara et al. (1987)]
  Yeast [Olson et al. (1986), Riles et al. (1993)]
  Nematode [Coulson et al. (1986)]

Early Physical Mapping of Human Chromosomes

- Strategies Analogous to those Used with *E. coli*, Yeast, and Nematode

  Applied to Several Human Chromosomes
  Cosmid Clones (e.g., Flow-Sorted Libraries)
  Restriction Map Construction and/or Fingerprint Analysis
  [e.g., Stallings et al. (1990)]

- Shift in Strategies with the Development of YACs

- Distinguishing Features of YACs: No Ability to Readily Purify Cloned DNA

- Modified Fingerprint-based Strategies Attempted with YACs
  [e.g., Bellanne-Chantelot et al. (1992)]

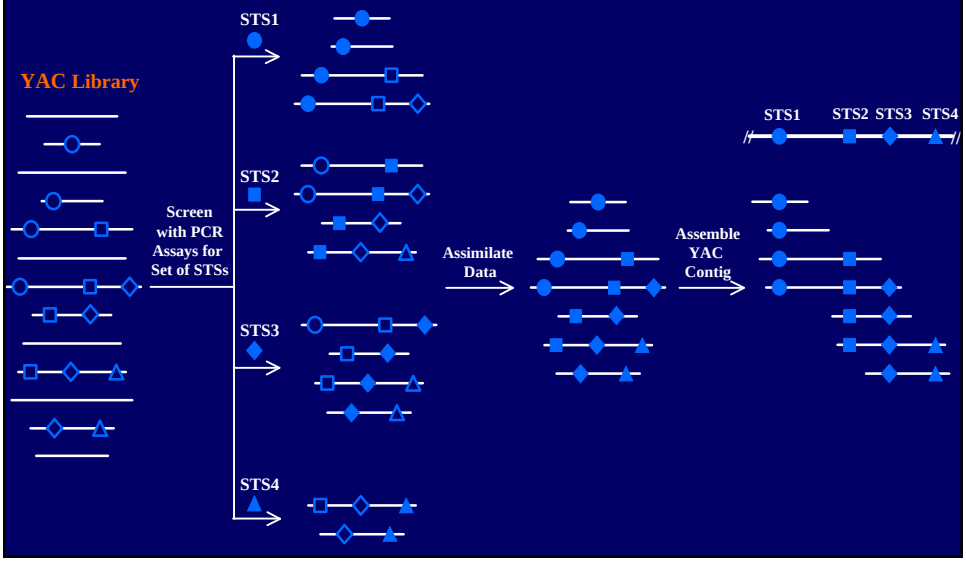
  1. Requires Gel-Transfer Hybridization
  2. Typically Uses Repetitive Element-Specific Probe(s)
     Establish YAC 'Fingerprint' → Infer Overlap(s) with Other YACs

- Development of PCR → Sequence-Tagged Sites (STSs)

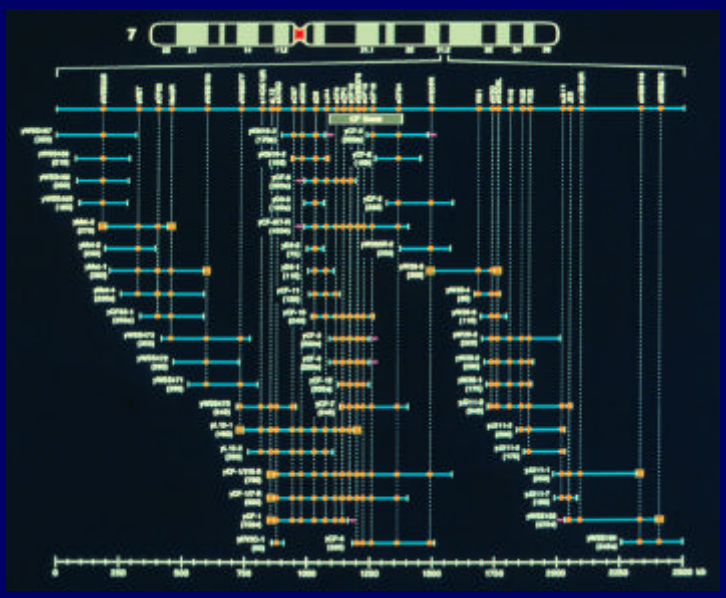
- 'Common Language' of STSs Proposed by Olson et al. (1989)

YAC-Based STS-Content Mapping

(E.D. Green & P. Green, PCR Meth. Applic. 1:77, 1991)



YAC-Based STS-Content Map



Green & Olson, *Science* 250:94-98, 1990

STSs as Landmarks

- Advantages of STSs as Landmarks

  Independent of the Mapping Resource (Clones, RH Panel)
  PCR-based (Sensitivity, Specificity, Automation)
  Electronic-based Nature of STSs
  Sequence-based Nature Facilitates Integration with Sequence

- General Review on STS-Content Mapping: Green and Green (1991)

- Programmatic Goal of U.S. Human Genome Project [Collins & Galas (1993)]

  100-kb Average Resolution STS Map of Human Genome
  Therefore, ~ 30,000 STSs for Human Genome

- STS Map as 'Intermediate Map' En Route to Sequencing

- Conceptual Similarity of STSs and Probes

Development of STSs

- Operational Definition of an STS

  1. Sequence that Can be Amplified by a PCR Assay
  2. Functionally is 'Unique' in the Genome

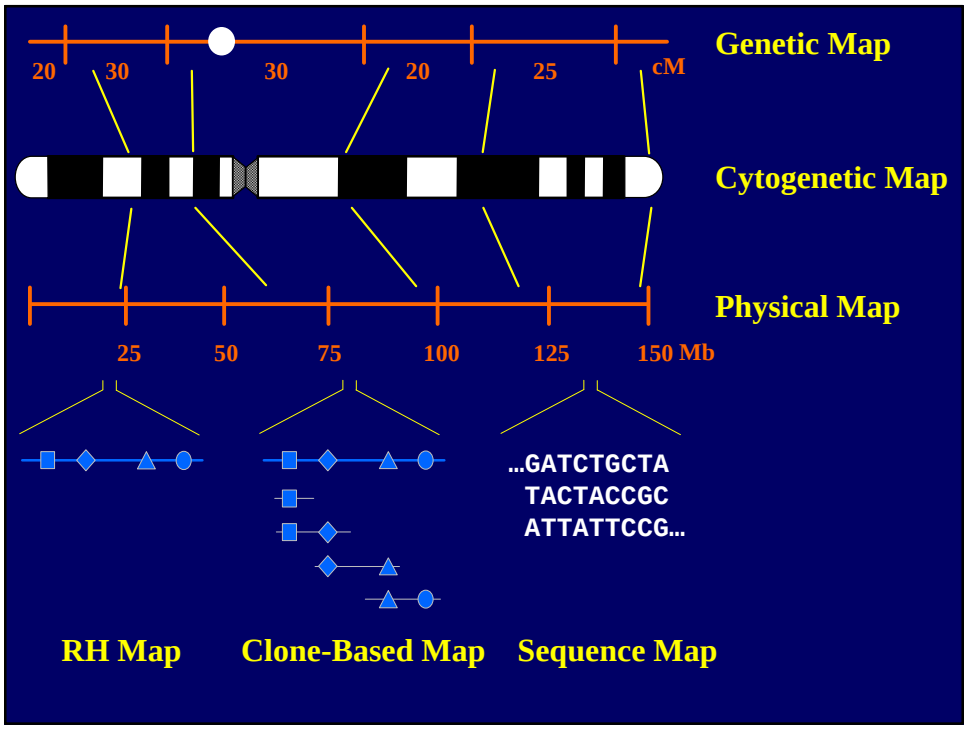
- DNA Sequence → Select Primers → Confirm Above Definition

- Generation of Sequence for Developing STSs (see Vollrath 1999)

  1. Non-Targeted (i.e., Genome-Wide)
  2. Targeted

- Targeted Approaches

  Specific Chromosomes
    Somatic Hybrid Cell Lines
    Flow Sorting
    Microdissection
  Genetic Markers (Microsatellites)
  Expressed Sequences [Genes, ESTs]



Map Integration

- **Rationale:**
  - Maximizes Utility of Maps
  - Assists in the Construction of Maps (i.e., Provides 'Cross-Checks')
- **Types of Map Integration:**
  - Physical & Genetic
  - Physical & Cytogenetic
  - Genetic & Cytogenetic

1st Generation Clone-based Physical Maps of the Human Genome

- Constructed with YACs
- Genome-Wide Efforts

  CEPH-Genethon
  Whitehead/MIT

- Chromosome-Specific Efforts

CEPH/Genethon YAC Map of Human Genome

·Bellanne-Chantelot et al. (1992), Cohen et al. (1993), Chumakov et al. (1995)

·Experimental Data Set

Hybridization-based Fingerprints
Hybridization Analysis (YAC x YAC) via *Alu*-PCR
Alu-PCR Hybridization Assignment of YACs to Chromosomes
FISH-Based Assignment of YACs to Chromosomes
***Assignment of Genethon Genetic Markers (STSs) to YACs

·Data Analysis

Complicated!!!
Suite of Programs to 'Disambiguate' the Data (*Quickmap*)
Heavy Reliance on Genethon Genetic Map for Contig Assembly
Predict 'Most Likely' Paths Among Overlapping Clones

·Map Highlights

225 Contigs Averaging 10 Mb, ~75% of Genome Covered
Potentially Useful for Positional Cloning Projects
Poor Scaffold for DNA Sequencing (Sparse STS Density)

·Data and Map Availability: www.cephb.fr/bio/ceph-genethon-map.html

Whitehead/MIT YAC Map of Human Genome

·Hudson et al. (1995)

·~25,000 STSs Mapped Relative to YACs and/or RH Panel and/or by Genetic Mapping

·Integrated Approach for Physical Mapping of STSs

1. YAC-based STS-Content Mapping
~11,000 STSs, CEPH Mega-YACs

2. RH Mapping of STSs
~15,000 STSs, GeneBridge 4 RH Panel

3. Genethon Genetic Maps
~5,300 STSs

·PCR Analysis: *Genomatron*

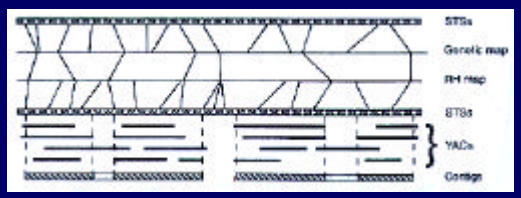
Massively-Parallel, Factory-Style Automation System
1,536 Position Arrays
~150,000 PCR Assays Per Run
>25,000,000 PCR Assays Total

Whitehead/MIT YAC Map of Human Genome

·Strategy for Map Construction

Genetic and RH Maps Provide Global Framework ('Top-Down Mapping')

YAC-based STS-Content Map Provides Local Ordering of STSs ('Bottom-Up Mapping')



Cross-Reference to Deduce an 'Integrated Map' of Each Chromosome

·Average STS Resolution: ~120 kb

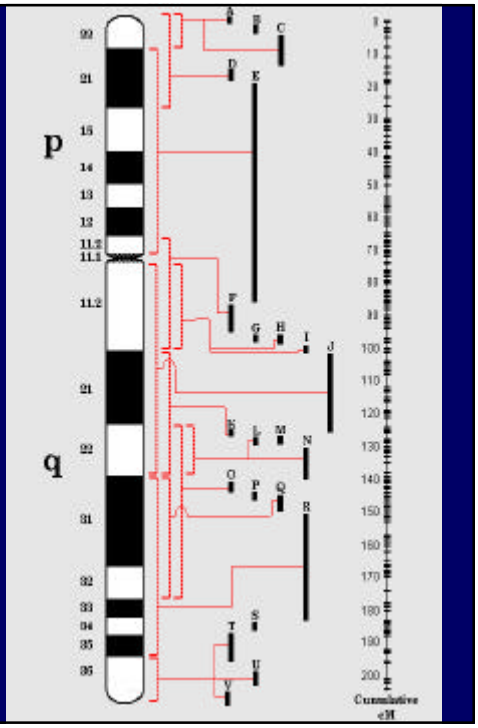
·Availability of Data and Maps: www-genome.wi.mit.edu

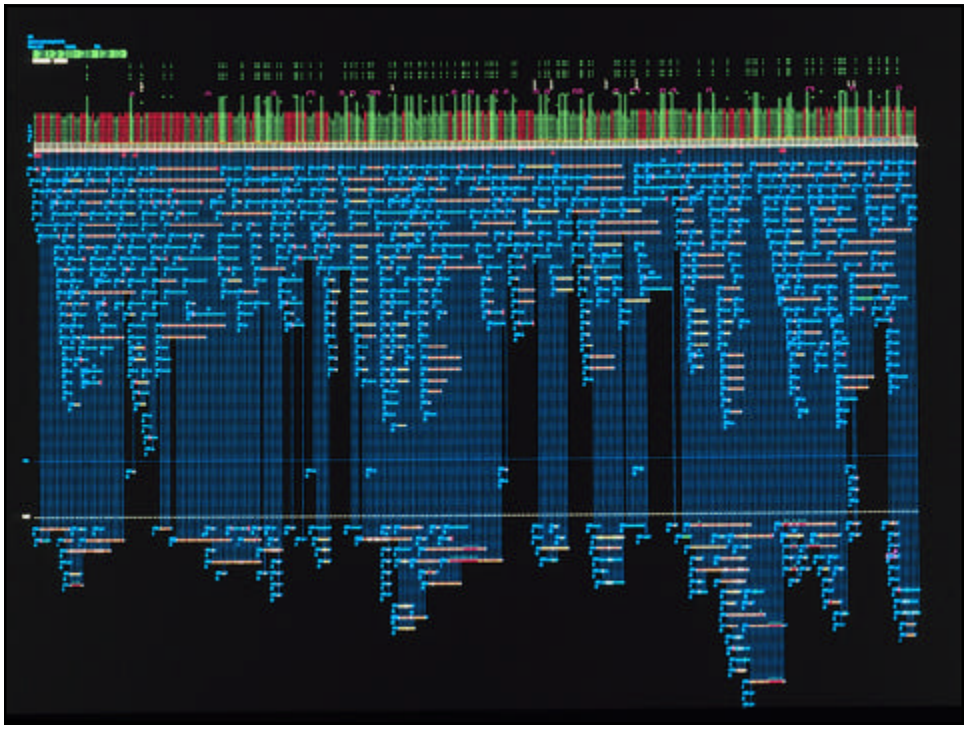
Chromosome-Specific YAC Maps of Human Genome

- ·Chromosome 3 Gemmill et al. (1995)
- ·Chromosome 4 Goold et al. (1993)
- ·Chromosome 7 Green et al. (1994, 1995), Bouffard et al. (1997)
- ·Chromosome 10 Genome Therapeutics, Unpublished
- ·Chromosome 11 Smith et al. (1993), Quackenbush et al. (1995), Qin et al. (1996)
- ·Chromosome 12 Krauter et al. (1995)
- ·Chromosome 16 Doggett et al. (1995)
- ·Chromosome 19 Ashworth et al. (1995)
- ·Chromosome 21 Chumakov et al. (1992), Korenberg et al. (1995), Wang et al. (1999)
- ·Chromosome 22 Bell et al. (1995), Collins et al. (1995)
- ·Chromosome X Nagaraja et al. (1997)
- ·Chromosome Y Foote et al. (1992), Vollrath et al. (1992)

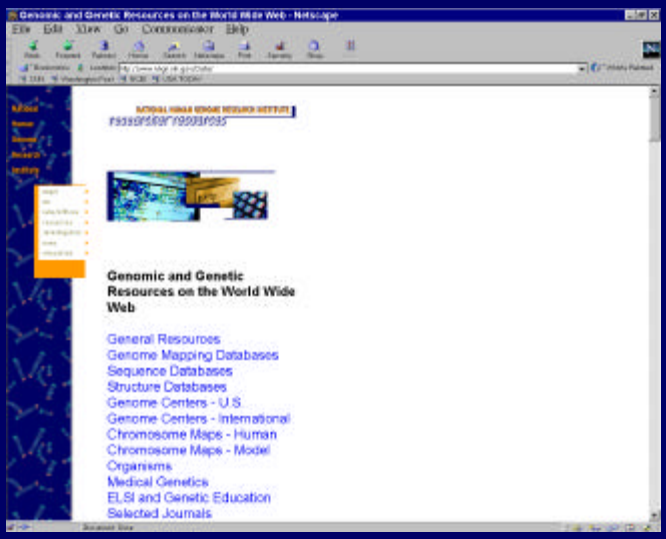
YAC-based Physical Map of Human Chromosome 7

Bouffard et al. (1997)

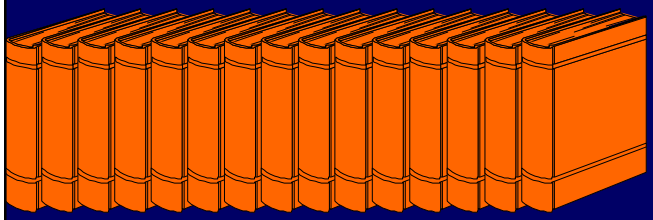




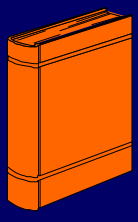
Guide to Web Sites with Physical Mapping Data



www.nhgri.nih.gov/Data



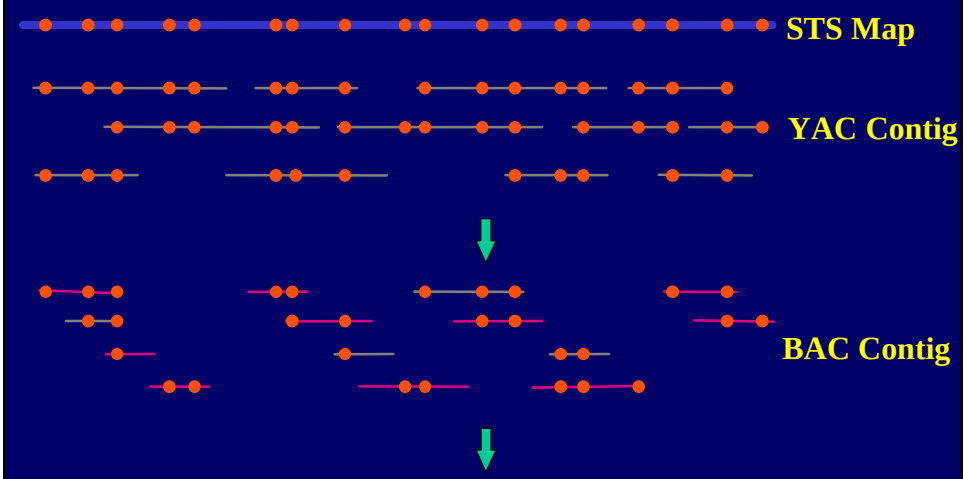
Genome
(~3000 Mb)



Chromosome
(~130 Mb)

BAC
(~0.1-0.2 Mb)

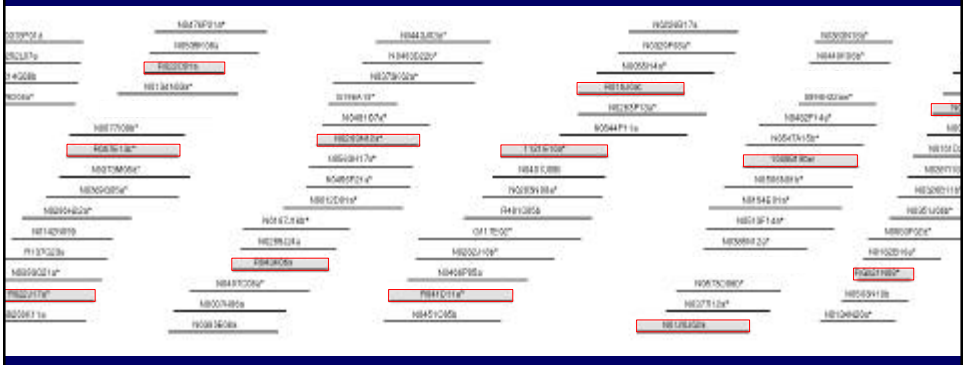
1st to 2nd Generation Maps of the Human Genome



Sequence Minimally Overlapping Clones

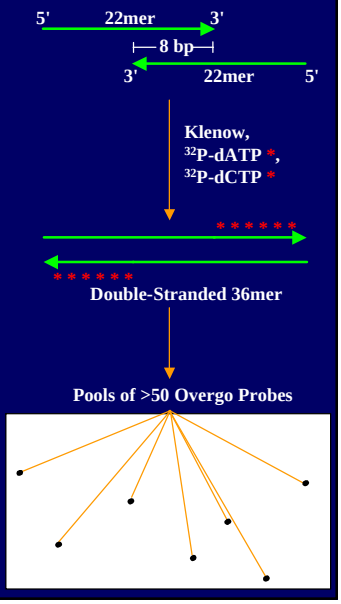
2nd Generation Physical Maps: 'Sequence-Ready Maps'

Marra et al. (1997), Gregory et al. (1997), Wong et al. (1997)

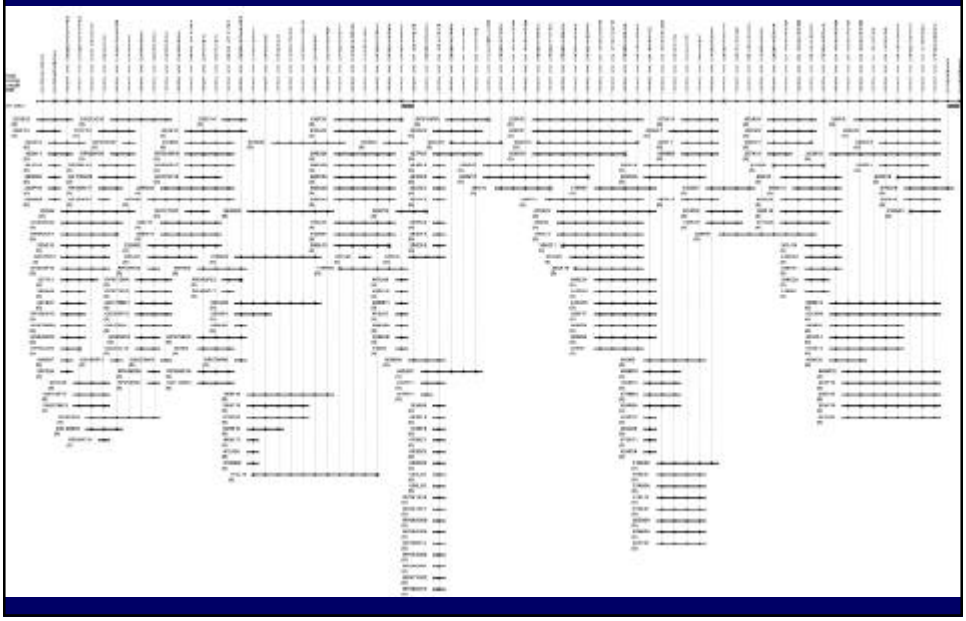


‘Overgo’ Hybridization Probes

- Pair of ~22mer Oligonucleotide Primers with 8-bp Overlap
- Primer Extension with Klenow and both ^{32}P-dATP and ^{32}P-dCTP
- Low Background Allows Pooling of Multiple Overgo Probes



Probe-Content BAC Contig Map



Restriction Enzyme Digest-based Fingerprint Analysis

- BAC DNA Purification in 96-Well Format
- HindIII Digestion
- 1% Agarose Gel

>20 kb

~300 bp

Marra et al., 1997

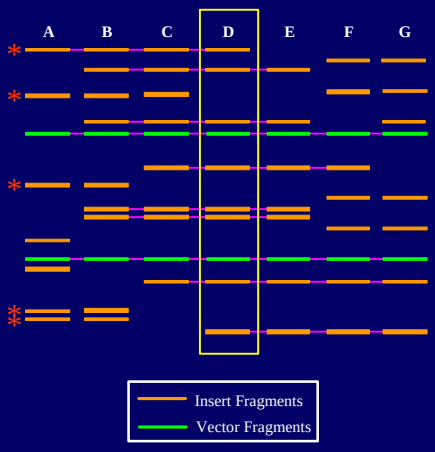
Restriction Enzyme Digest-based Fingerprint Analysis

Restriction Enzyme-Digested BAC DNA

Identify Overlapping Clones by the Presence of Common Restriction Fragments

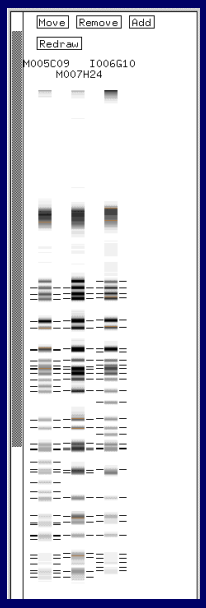
Process Repeated Iteratively to Assemble a BAC Contig

A Clone Selected for Sequencing Must Have All Restriction Fragments Accounted for in Overlapping Clones



Marra et al., 1997

IMAGE



FPC

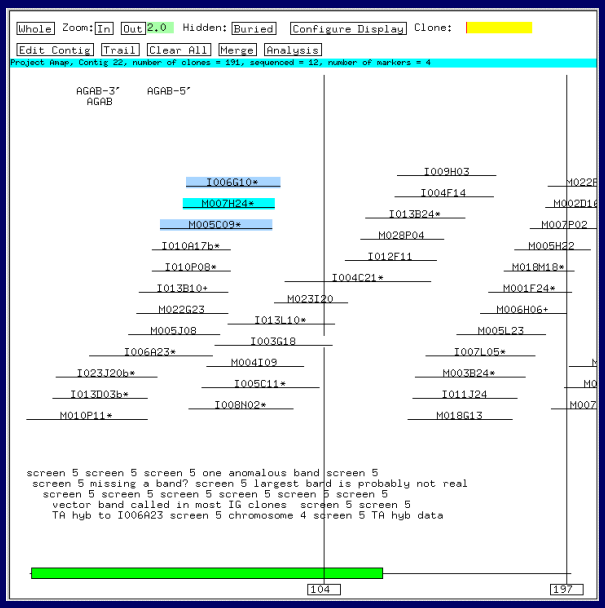
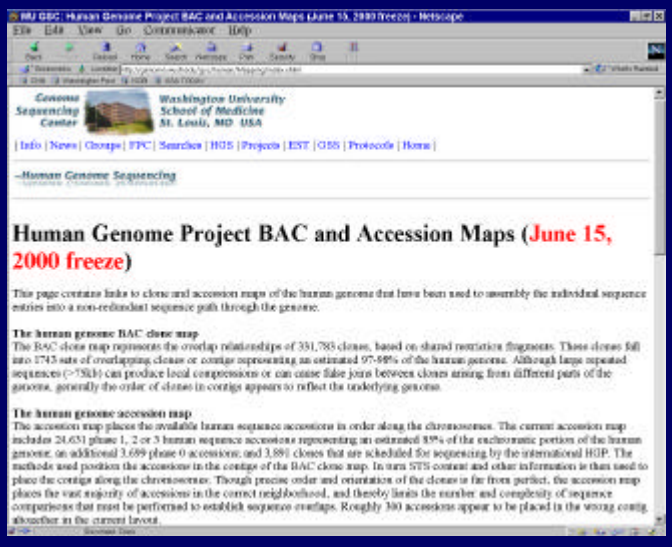


IMAGE and FPC: www.sanger.ac.uk/Software

BAC-based Physical Map of Human Genome



genome.wustl.edu/gsc/human/Mapping/index.shtml

BAC-based Physical Map of Mouse Genome

www.bcgsc.bc.ca/projects/mouse_mapping

Future Prospects: Physical Mapping of Other Vertebrates

Rapidly changing strategies with increases in sequencing capabilities…

Mouse

- **www.bcgsc.bc.ca/projects/mouse_mapping**
- **www.informatics.jax.org**
- **www-genome.wi.mit.edu**

Rat

- **www.informatics.jax.org/rat**

Zebrafish

- **zfin.org/ZFIN**

Others (non-human primates, dog, cat, cow, pig, etc.)???